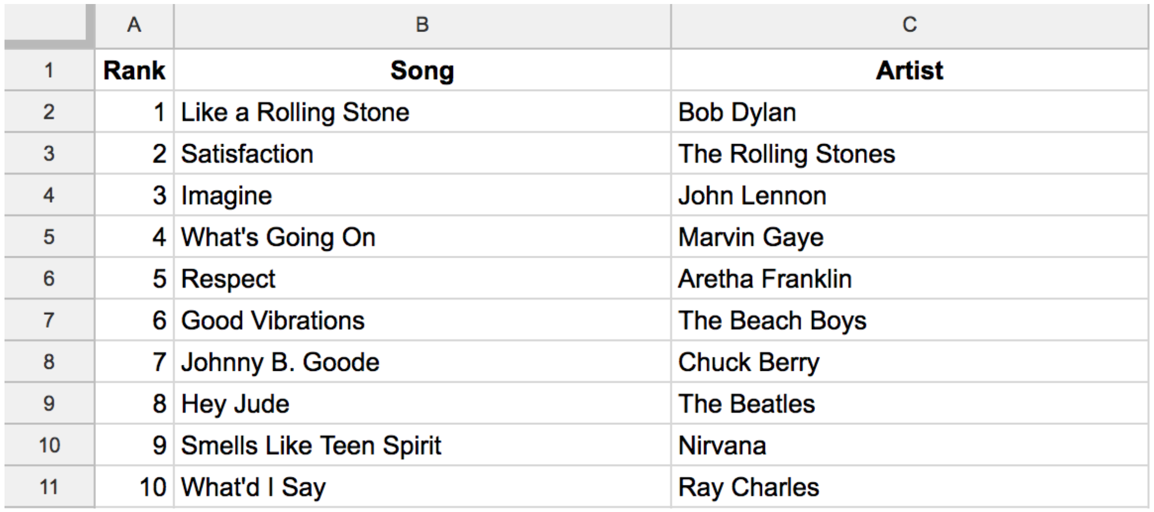
Compsci 101
Sorting, CSV



	A	B	C
1	**Rank**	**Song**	**Artist**
2	1	Like a Rolling Stone	Bob Dylan
3	2	Satisfaction	The Rolling Stones
4	3	Imagine	John Lennon
5	4	What's Going On	Marvin Gaye
6	5	Respect	Aretha Franklin
7	6	Good Vibrations	The Beach Boys
8	7	Johnny B. Goode	Chuck Berry
9	8	Hey Jude	The Beatles
10	9	Smells Like Teen Spirit	Nirvana
11	10	What'd I Say	Ray Charles

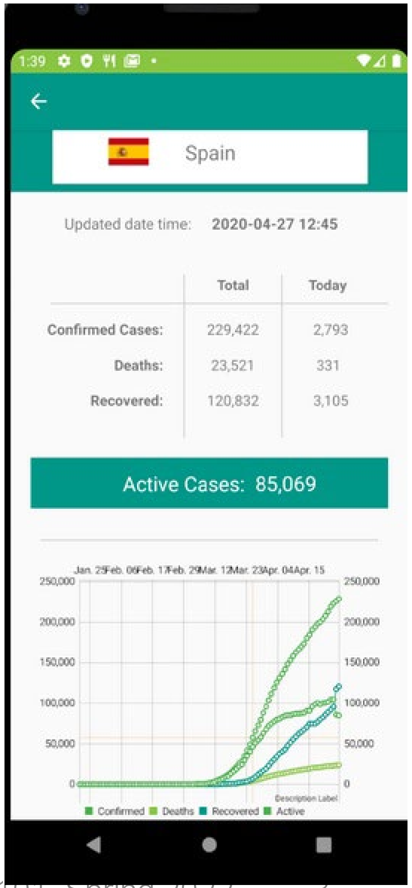
Susan Rodger
March 24, 2022

`R` is for …

- **Random**
  - **`.choice`**, **`.shuffle`**, **`.seed`**, **`.randint`**
- **R**
  - Programming language of choice in stats
- **Refactoring**
  - A way to rename your variable, function name

Esther Brown

- Duke Alum 2020, IDM CS/Cult. Anth.
- Harvard MS Data Sci
- Starting PhD in CS at Harvard!
- At Duke, as Senior did I.S. creating five Apps
  - Covid tracker
  - Movie App

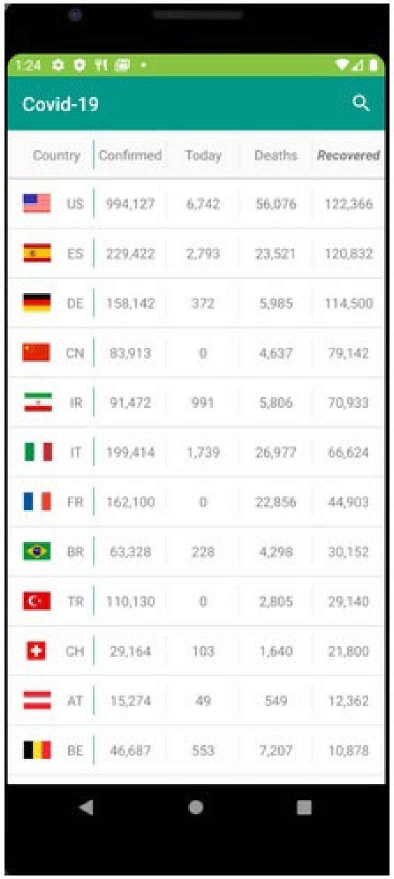


Spain

Updated date time: 2020-04-27 12:45

	Total	Today
Confirmed Cases:	229,422	2,793
Deaths:	23,521	331
Recovered:	120,832	3,105

Active Cases: 85,069



Covid-19

Country	Confirmed	Today	Deaths	Recovered
US	994,127	6,742	56,076	122,366
ES	229,422	2,793	23,521	120,832
DE	158,142	372	5,985	114,500
CN	83,913	0	4,637	79,142
IR	91,472	991	5,806	70,933
IT	199,414	1,739	26,977	66,624
FR	162,100	0	22,856	44,903
BR	63,328	228	4,298	30,152
TR	110,130	0	2,805	29,140
CH	29,164	103	1,640	21,800
AT	15,274	49	549	12,362
BE	46,687	553	7,207	10,878

Announcements

- APT 5 due today!
- Assignment 5 due Tue, March 29
- APT-6 out today, due Thur, March 31

- Lab 9 Friday
  - There is NO prelab!

- Reading and Sakai Quizzes due next week

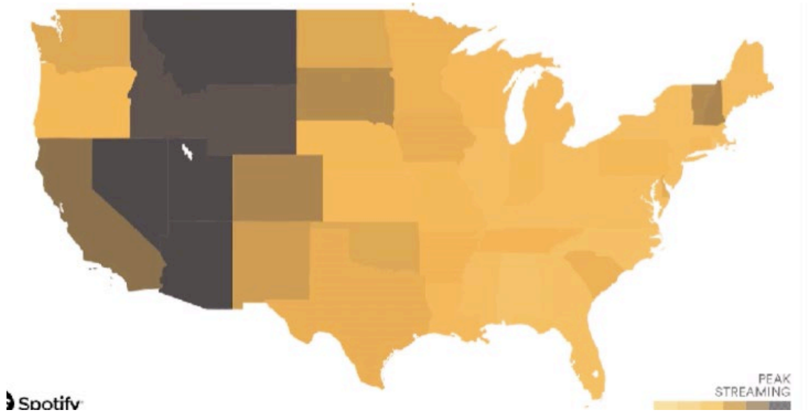
PFTD

- Sorting
  - Sorting using standard Python APIs
- CSV Library
  - How to read data using standard Python APIs
- Lambda
  - Language construct to make sorting simpler (next week)

Song: Total Eclipse of the Heart, Bonnie Tyler
https://www.youtube.com/watch?v=lcOxhH8N3Bo

Why Sort Data?

- **Help understand data**
  - Great American Eclipse, August 21, 2017
  - http://bit.ly/spotify-eclipse-cnet
  - Spotify tracked the playing of the song



Why Sort Data?

- **Every field needs to visualize and understand data**
  - Sorting helps with this from movies to policy to sports to location of infections to

https://www.esri.com/arcgis-blog/products/apps/local-government/how-your-gis-department-can-respond-to-covid-19/

How your GIS department can respond to COVID-19

Local Government
March 09, 2020

Mike Schoelen

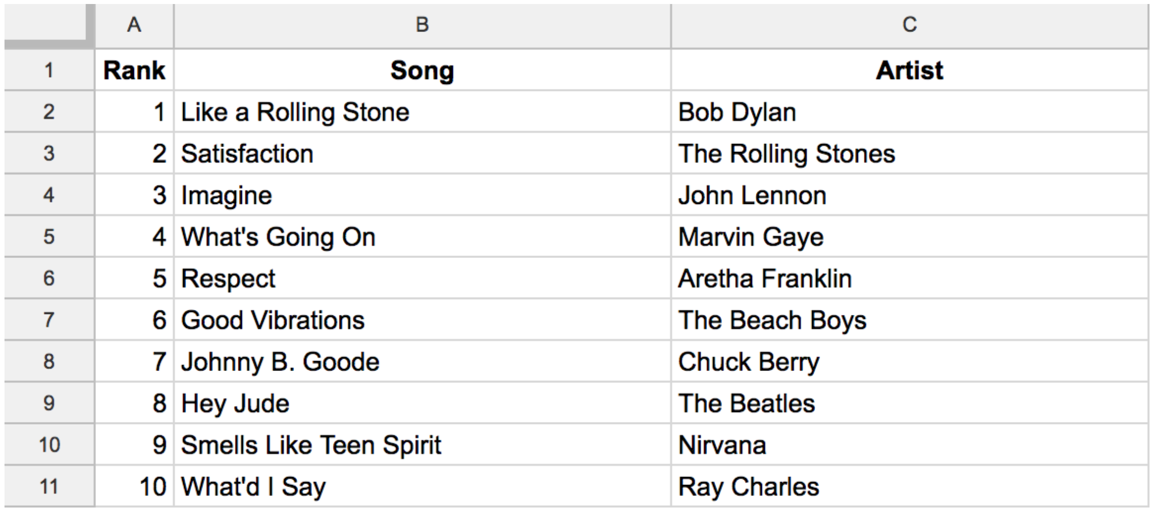
A staggering wealth of geospatial information has emerged regarding the COVID-19 outbreak. Dashboards, near real-time services, and GitHub repositories have built the foundation for an extraordinarily transparent response effort.

How To Sort: Algorithms

- Does scale matter? It depends!
- You're playing Spades, Hearts, Bridge, Go-Fish
  - How you sort doesn't really matter, but whether you sort makes play more efficient? Better?
- Many ways to sort
  - Bubblesort, Insertion sort, Selection sort
  - Quicksort, Mergesort, Timsort, Bogo sort
  - Python uses Timsort

WOTO-1 Popular Music http://bit.ly/101s22-0324-1

- Make a copy of this spreadsheet:
  - http://bit.ly/101s22-0324-data



	A	B	C
1	**Rank**	**Song**	**Artist**
2	1	Like a Rolling Stone	Bob Dylan
3	2	Satisfaction	The Rolling Stones
4	3	Imagine	John Lennon
5	4	What's Going On	Marvin Gaye
6	5	Respect	Aretha Franklin
7	6	Good Vibrations	The Beach Boys
8	7	Johnny B. Goode	Chuck Berry
9	8	Hey Jude	The Beatles
10	9	Smells Like Teen Spirit	Nirvana
11	10	What'd I Say	Ray Charles

Solve a Larger Problem

- Suppose I were to give you the top 1000 artists
  - Top 1,000 songs, find top 10 artists
  - How many songs per artist?

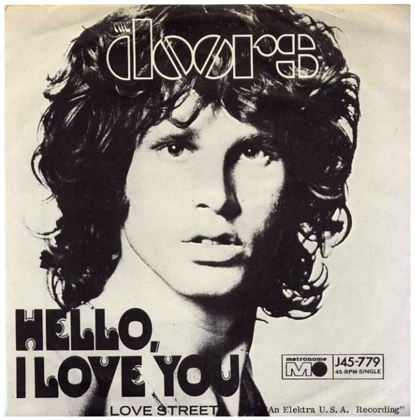
Scale

- As the size of the problem grows we want …
  - The algorithm to still work and be fast!
  - What to do?

- Search example
  - Google search results work
  - SoundHound/Shazam results work
  - ContentID on YouTube results work

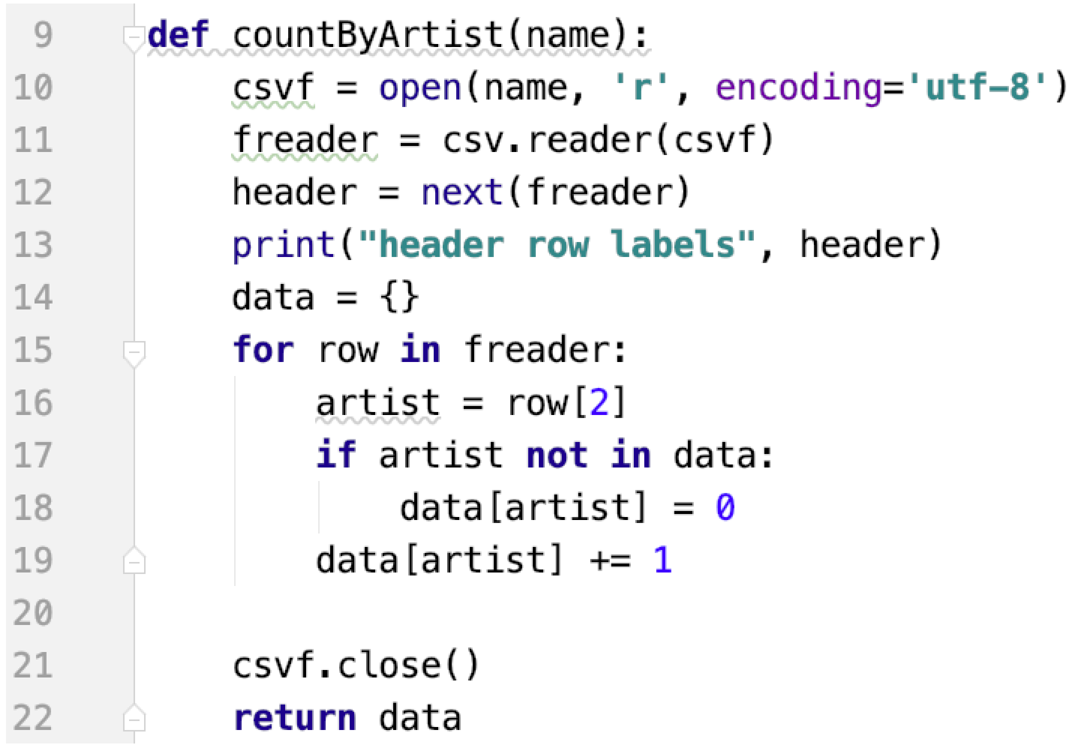
Python to the Rescue

- Using `.sort(…)`, `sorted(…)`, and `lambda`
- Using CSV library and its API
  - CSV – Comma Separated Values
- Why use the CSV library?
  - How to handle the song “Hello, I Love You”?
  - Row 166 in spreadsheet



Hits by Artists: SongReader.py

- What is returned by this function?
  - details of csv: **`next`** and no **`split`** and …



```python
def countByArtist(name):
    csvf = open(name, 'r', encoding='utf-8')
    freader = csv.reader(csvf)
    header = next(freader)
    print("header row labels", header)
    data = {}
    for row in freader:
        artist = row[2]
        if artist not in data:
            data[artist] = 0
        data[artist] += 1

    csvf.close()
    return data
```

WOTO-2 countByArtist
http://bit.ly/101s22-0324-2

Two APIs: CSV and Sorting

- CSV Library to read and process data
  - Comma-separated, but can by ":" separated, or any character as we'll see later

- Similar to reading a file – returned by open
  - Iterable is returned by **`csv.reader`**
  - The **`next`** function advances iterable
  - Don't call **`split`**, we can access by index
    - Also by header-row label with **`csv.dictreader`**

CSV API

- freader = csv.reader(file) – returns an iterable
  - Every line from the file in a form ready for you
- line = next(freader)
  - Gives you next row as list of strings
- for row in freader:
  - Gives you the rest of rows as list of strings

What does this do? freader an iterable Where name is a filename

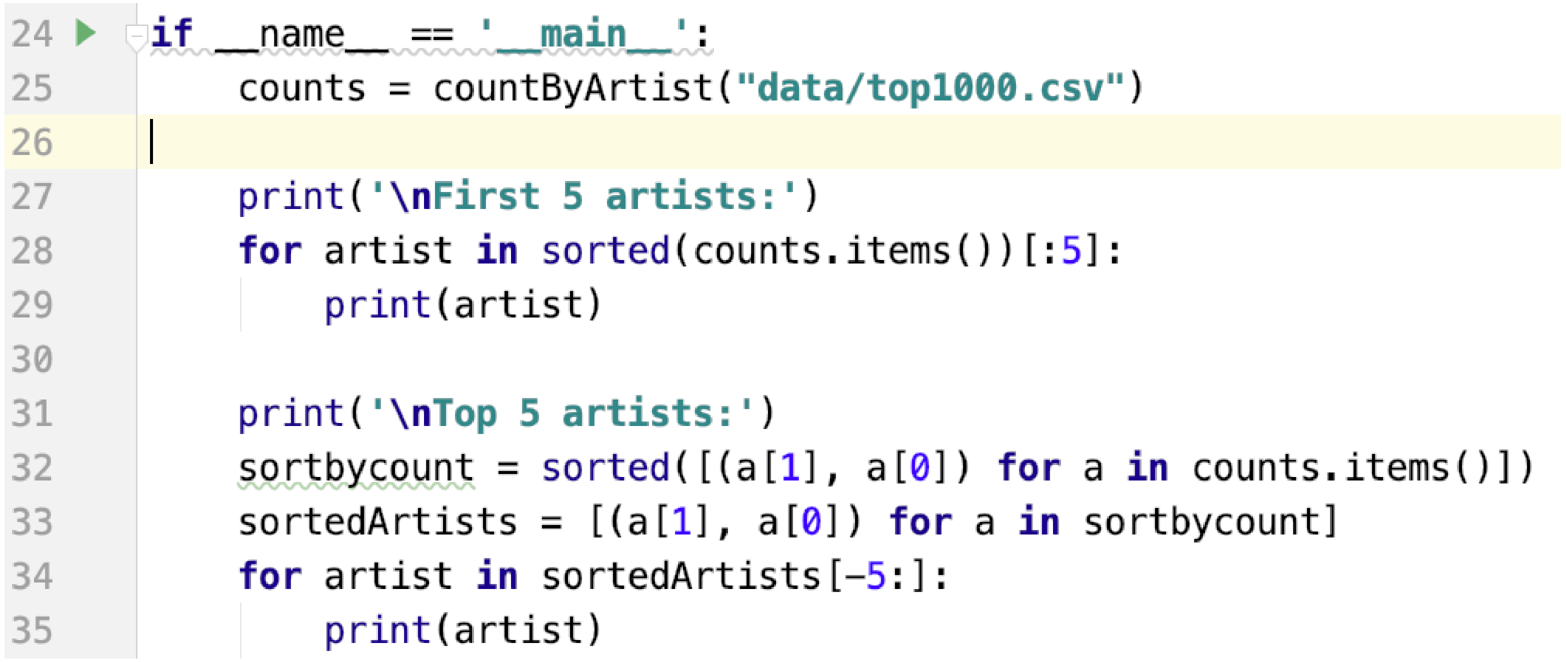
```
csvf = open(name, 'r', encoding='utf-8')
freader = csv.reader(csvf)
print("freader", freader)
header = next(freader)
print("header", header)
for row in freader:
    print("row", row)
```

What if you call **next** one extra time? Where name is a filename

```
csvf = open(name, 'r', encoding='utf-8')
freader = csv.reader(csvf)
print("freader", freader)
header = next(freader)
print("header", header)
nextline = next(freader)
print("next", nextline)
for row in freader:
   print("row", row)
```

Sorting to Print/Visualize

- Dictionary is ('Beatles', 51) tuples
  - But tuples not in order, so we must …

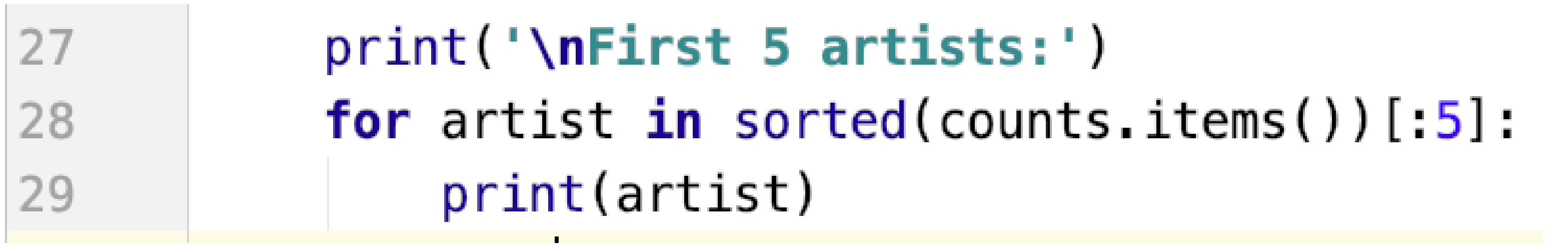


```
if __name__ == '__main__':
    counts = countByArtist("data/top1000.csv")

    print('\nFirst 5 artists:')
    for artist in sorted(counts.items())[:5]:
        print(artist)

    print('\nTop 5 artists:')
    sortbycount = sorted([(a[1], a[0]) for a in counts.items()])
    sortedArtists = [(a[1], a[0]) for a in sortbycount]
    for artist in sortedArtists[-5:]:
        print(artist)
```

WOTO-3 Calling countByArtist
http://bit.ly/101s22-0324-3

Sorting API and Sorting Concepts

- What is **`counts.items()`** – how is it sorted?

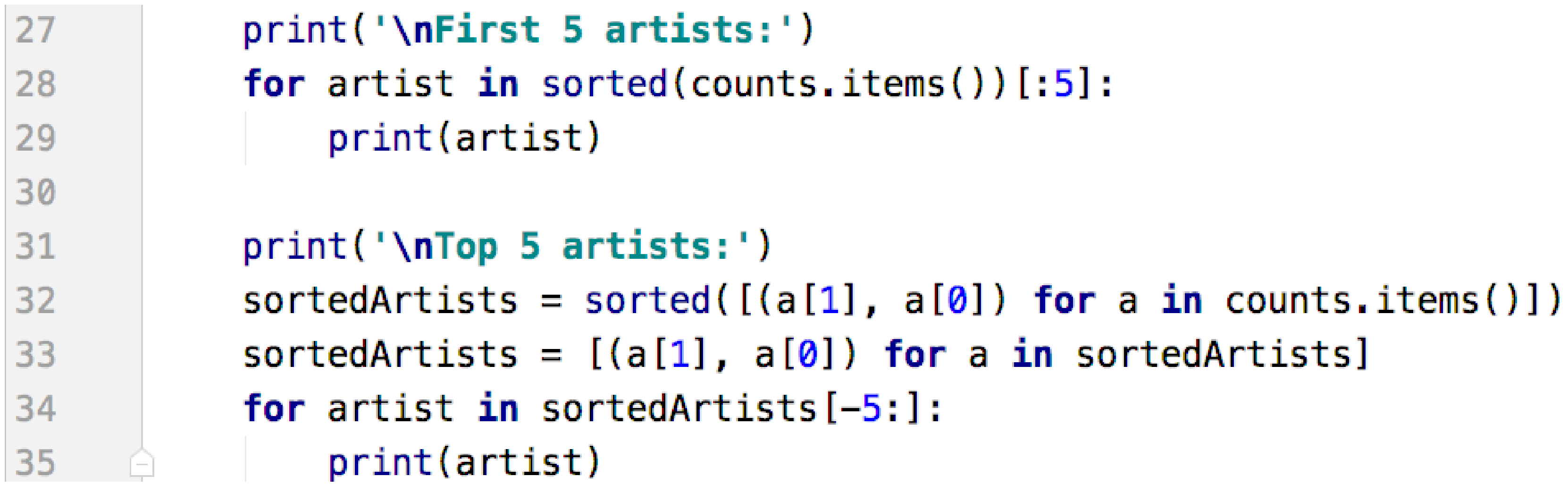


```
    print('\nFirst 5 artists:')
    for artist in sorted(counts.items())[:5]:
        print(artist)
```

- What does sorted return?
  - A list, you can slice a list, look for clues!
  - What can be sorted? A sequence
  - **`sorted(counts.items())`**

Sorting by Number of Songs

- Sort by first value vs sort by second value
  - Need to put sequence back to original format



```python
    print('\nFirst 5 artists:')
    for artist in sorted(counts.items())[:5]:
        print(artist)

    print('\nTop 5 artists:')
    sortedArtists = sorted([(a[1], a[0]) for a in counts.items()])
    sortedArtists = [(a[1], a[0]) for a in sortedArtists]
    for artist in sortedArtists[-5:]:
        print(artist)
```

Python Sorting API

- We'll use both **`sorted()`** and **`.sort()`** API
  - How to call, what options are
  - How to sort on several criteria

- Creating a new list, modifying existing list
  - **`sorted(..)`** creates list from .. Iterable
  - **`x.sort()`** modifies the list x, no return value!

API to change sorting

- In SongReader.py we changed order of tuples to change sorting order
  - Then we sliced the end to get "top" songs

- Can supply a function to compare elements
  - Function return value used to sort, key=function
  - Change order: reverse=True

Sorting Examples

- Use key=function argument and reverse=True
  - What if we want to write our own function?

```
a = [‘red’, ‘orange’, ‘green’, ‘blue’, ‘indigo’, ‘violet’]
print(sorted(a))

print(sorted(a, key=len))

print(sorted(a, key=len, reverse=True))
```

WOTO-4 Sorting
http://bit.ly/101s22-0324-4